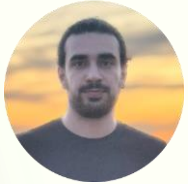
NEURAL INFORMATION PROCESSING SYSTEMS

Learning Truncated Causal History Model for Video Restoration

♣Amirhosein Ghasemabadi[1,2], ♣Muhammad Kamran Janjua[2], Mohammad Salameh[2], Di Niu[1]

1 ECE Department, University of Alberta
2 Huawei Technologies, Canada
♣ indicates equal contribution




HUAWEI

UNIVERSITY OF ALBERTA

Introduction to Video Restoration

'If you want to understand today, you have to search yesterday.'
– Pearl S. Buck

Video Restoration:

- Aims to improve low-quality videos affected by factors such as:
  - ❖ Motion Blur
  - ❖ Weather
  - ❖ Noise
  - ❖ Camera Sensors or Acquisition Procedure.

Challenges:

1. Effective information fusion across multiple frames
2. Handling non-uniform motion between frames.

Limitations of Existing Methods

Parallel Methods

- Process multiple frames simultaneously
- Multiple branches for feature extraction and reconstruction of each frame or set of frames
- Mix features mid-process to improve context
- High memory and computational cost

Recurrent Methods

- Process frames sequentially
- Some designs use auto-regression, feeding the output from the previous timestep as input along with the current degraded frame
- Lower memory use but prone to error accumulation
- Slower training due to limited parallelization

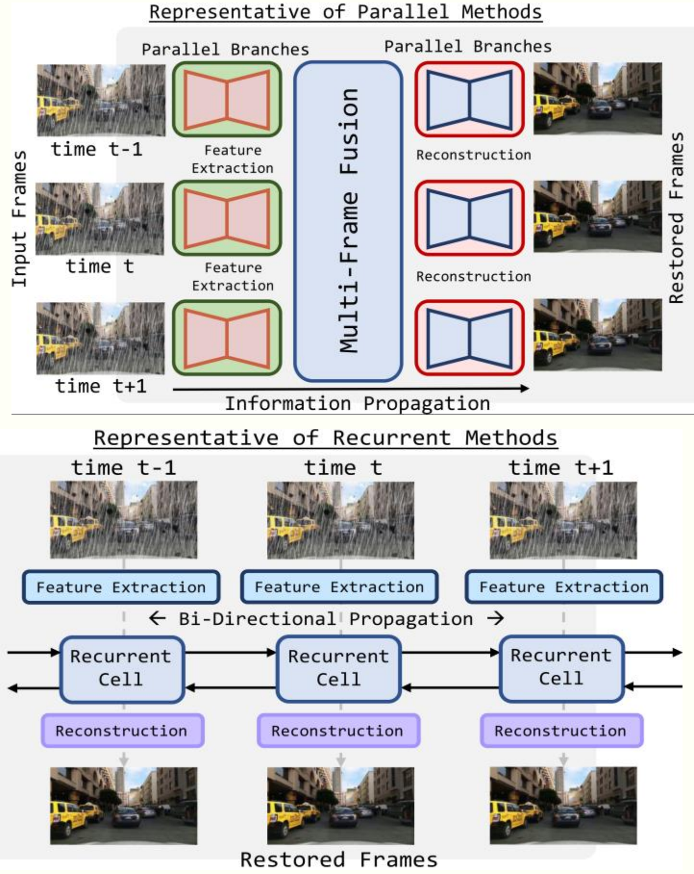


Figure 1: Parallel vs Recurrent Methods.

Overview

TURTLE:

- A video restoration framework designed to improve compute efficiency and quality.

Key Features

Online Video Processing:

- TURTLE processes each frame independently within the encoder.

Truncated Causal History:

- Uses a limited set of past frames to save memory.

Causal History Model (CHM) :

- Models the trajectory by summarizing the evolving frames into history states.
- Borrows information from the history states to compensate the input frame for motion and re-weights the entire trajectory to accentuate necessary information.

Tasks:

- TURTLE achieves state-of-the-art results on seven restoration tasks, including desnowing, deraining, super-resolution, and deblurring.

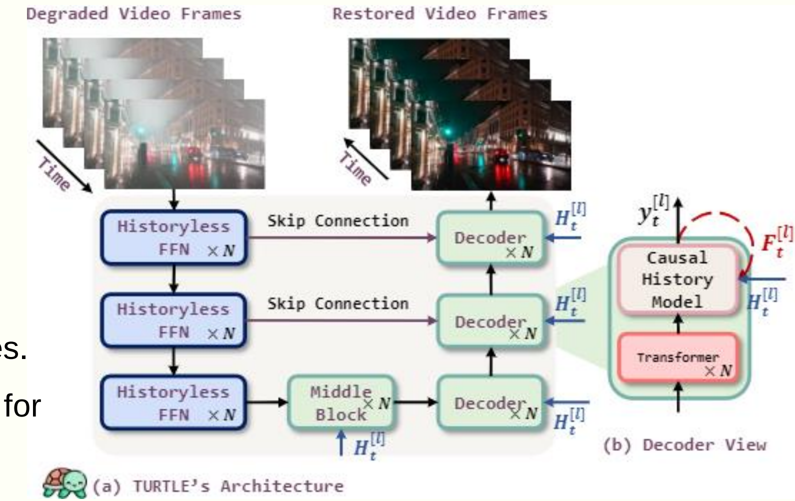


Figure 2: TURTLE Architecture.

Architecture details

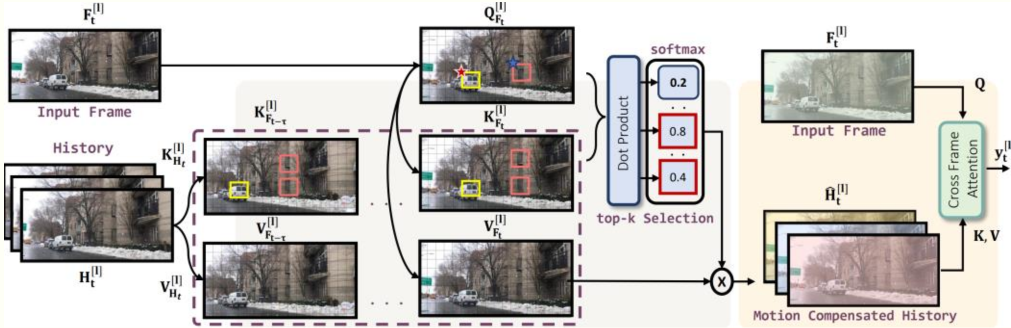


Figure 3: Causal History Model Visualized.

Encoder:

- Processes each frame independently, without relying on neighboring frames in the video.

Decoder:

- Uses aligned features from previously restored frames through the **Causal History Model (CHM)**.

CHM Function:

- **History Summarization:** CHM extends the state-space modeling paradigm to video processing and maintains an evolving state that summarizes the history of the frame.

- **Motion Compensation:** CHM aligns history states with the input frame through attention mechanism limited to topk most similar regions in the history.

- **Feature Re-weighting:** Prioritizes relevant features over time by re-weighting the entire trajectory, and the irrelevant information is suppressed.

Is CHM Necessary?

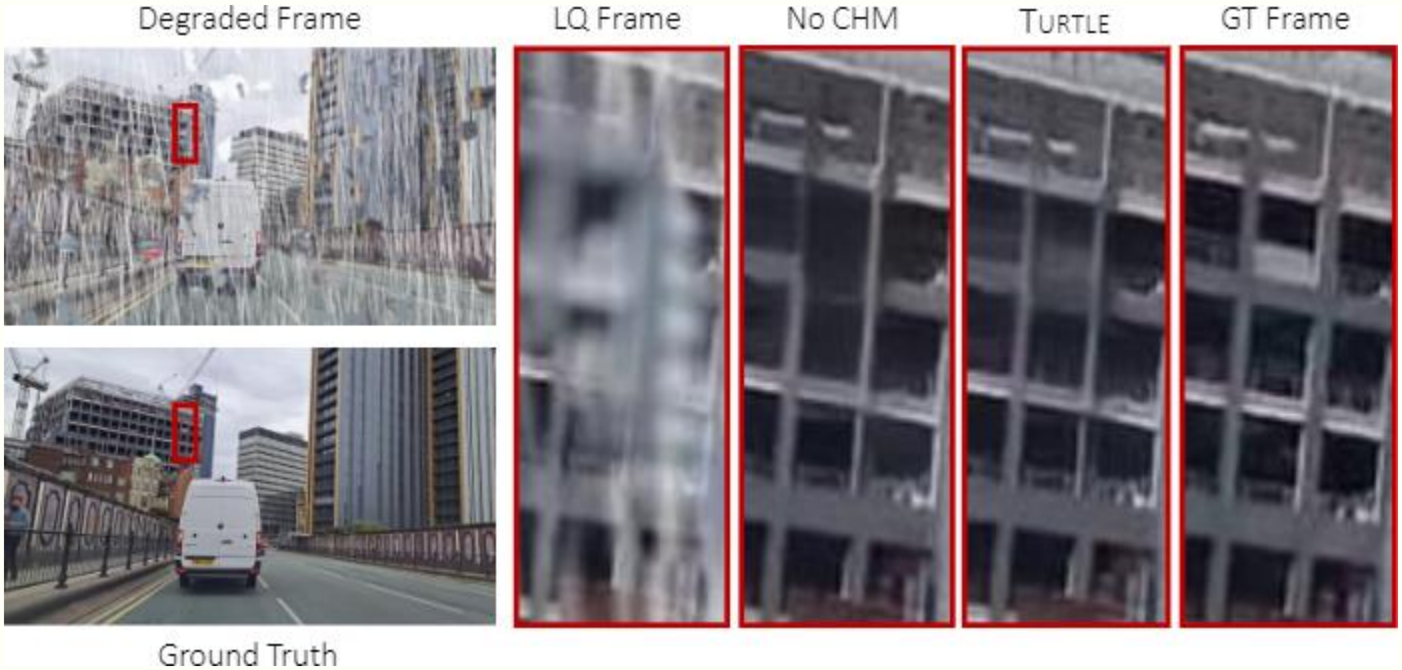


Figure 4: Is CHM Necessary?

Technical Features - 1

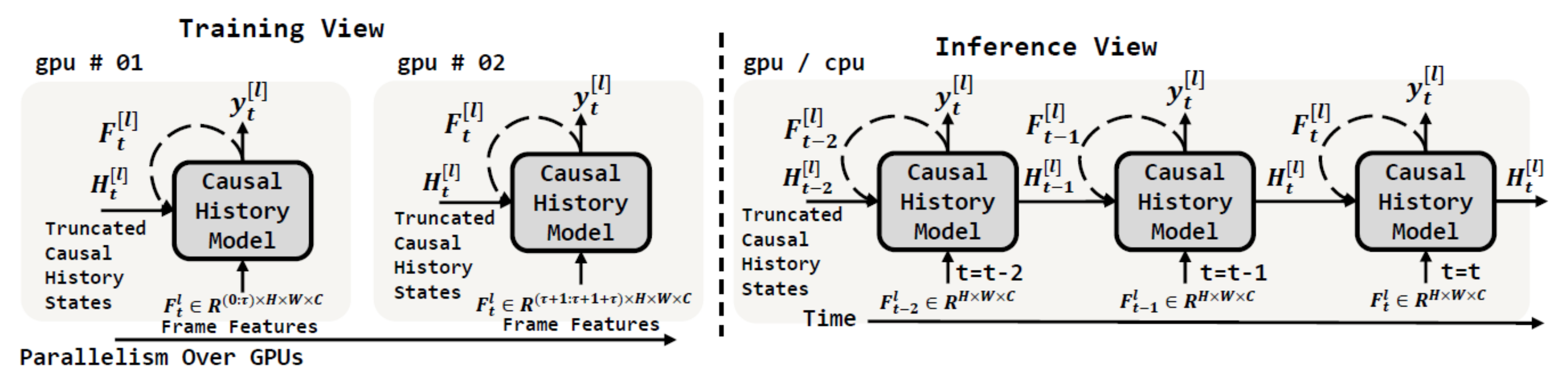


Figure 5: Stateless vs Stateful Configuration.

Configurable Mode:

- TURTLE can either be stateful or stateless.

Training:

- In training, TURTLE uses parallelism by dividing videos into clips, and minimizes recurrence.

Inference:

- In inference, TURTLE resorts to stateful configuration and implicitly maintains the entire trajectory to leverage longer temporal context for restoration.

Technical Features - 2

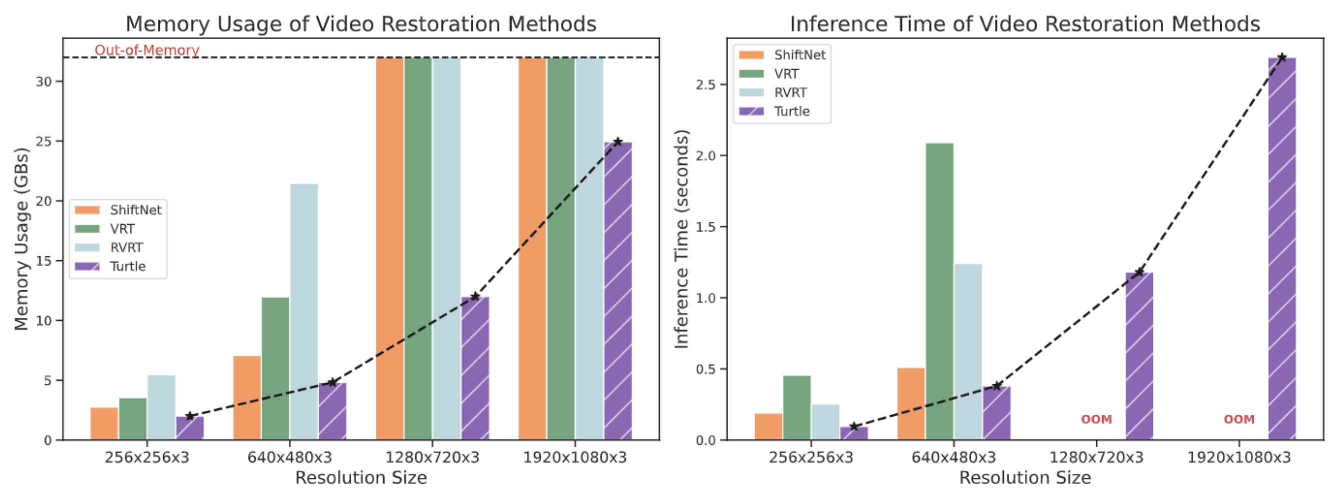


Figure 6: Run-Time and Memory Profile of TURTLE.

Efficiency:

- TURTLE reuses features from a limited set of previous frames, trading off compute for memory.
- Limits the motion compensation to *topk* most similar regions in the history.

Performance:

- Can process 1080p videos on a single consumer-grade 32 GB GPU, while many state-of-the-art methods encounter Out-of-Memory errors.

Experiments

Table 1: **Night Video Deraining Results.**

Method	PSNR↑	SSIM↑
FDM [22]	23.49	0.7657
DSTFM [46]	17.82	0.6486
WeatherDiff [43]	20.98	0.6697
RMFD [75]	16.18	0.6402
DLF [74]	15.17	0.6307
HRIR [31]	16.83	0.6481
MetaRain (Meta) [47]	23.49	0.7171
MetaRain (Scrt) [47]	22.21	0.6723
NightRain [35]	26.73	0.8647
TURTLE	**29.26**	**0.9250**

Table 2: **Video Desnowing Results.**

Method	PSNR↑	SSIM↑
TransWeather [65]	23.11	0.8543
SnowFormer [12]	24.01	0.8939
S2VD [78]	22.95	0.8590
RDDNet [68]	22.97	0.8742
EDVR [69]	17.93	0.5790
BasicVSR [6]	22.46	0.8473
IconVSR [6]	22.35	0.8482
BasicVSR++ [7]	22.64	0.8618
RVRT [33]	20.90	0.7974
SVDNet [10]	25.06	0.9210
TURTLE	**26.02**	**0.9230**

Table 6: **Blind Video Denoising Results.** Quantitative results on blind video denoising task in terms of distortion metrics, PSNR and SSIM, on two datasets DAVIS [48], and Set8 [61].

Methods	DAVIS		Set8	
	$\sigma = 30$	$\sigma = 50$	$\sigma = 30$	$\sigma = 50$
VLNB [1]	33.73	31.13	31.74	29.24
FastDVDNet [62]	34.04	31.86	31.60	29.42
DVDNet [61]	34.08	31.85	31.79	29.56
UDVD [55]	33.92	31.70	32.01	29.89
ReMoNet [72]	33.93	31.65	31.59	29.44
BSVD-32 [49]	34.46	32.25	31.71	29.62
BSVD-64 [49]	**34.91**	**32.72**	32.02	29.95
TURTLE	34.48	32.38	**32.22**	**30.29**

Table 3: **Real-World Video Deblurring.** Quantitative results (PSNR, and SSIM) on the 3ms-24ms BSD dataset [83] comparing state-of-the-art methods.

Method	PSNR↑	SSIM↑
STRCNN [24]	29.42	0.893
DBN [58]	31.21	0.922
SRN [60]	28.92	0.882
IFI-RNN [42]	30.89	0.917
STFAN [86]	29.47	0.872
CDVD-TSP [44]	31.58	0.926
PVDNet [57]	31.35	0.923
ESTRNN [83]	31.39	0.926
TURTLE	**33.58**	**0.954**

Table 4: **Synthetic Video Deblurring Results.** Quantitative results (PSNR, and SSIM) on the GoPro dataset [41] comparing state-of-the-art methods.

Method	PSNR↑	SSIM↑
IFI-RNN [42]	31.05	0.9110
ESTRNN [82]	31.07	0.9023
EDVR [69]	31.54	0.9260
TSP [44]	31.67	0.9280
GSTA [59]	32.10	0.9600
FGST [36]	32.90	0.9610
BasicVSR++ [7]	34.01	0.9520
DSTNet [45]	34.16	0.9679
TURTLE	**34.50**	**0.9720**

Table 5: **Video Raindrop and Rain Streak Removal.** Quantitative results (PSNR, and SSIM) on the VRDS dataset [71] comparing state-of-the-art methods.

Method	PSNR↑	SSIM↑
S2VD [78]	18.95	0.6630
EDVR [69]	19.19	0.6363
BasicVSR [6]	28.35	0.8990
VRT [34]	27.77	0.8856
TTVSR [37]	28.05	0.8998
RVRT [33]	28.24	0.8857
RDDNet [68]	28.38	0.9096
BasicVSR++ [7]	29.75	0.9171
ViMPNet [71]	31.02	0.9283
TURTLE	**32.01**	**0.9590**

Table 7: 4× **Video Super Resolution.** Quantita tive results on video super resolution task in term of distortion metrics, PSNR and SSIM.

Method	PSNR↑	SSIM↑
TDAN [63]	23.07	0.7492
EDVR [69]	23.51	0.7611
BasicVSR [6]	23.38	0.7594
MANA [76]	23.15	0.7513
TTVSR [37]	23.60	0.7686
BasicVSR++ [7]	23.70	0.7713
EAVSR [67]	23.61	0.7618
EAVSR+ [67]	23.94	0.7726
TURTLE	**25.30**	**0.8272**

Figure 7: TURTLE Results on Video Restoration Tasks

Visual Results -1

Figure 8: Synthetic Video Deblurring Results.

Figure 9: Video Desnowing Results.

Visual Results -2

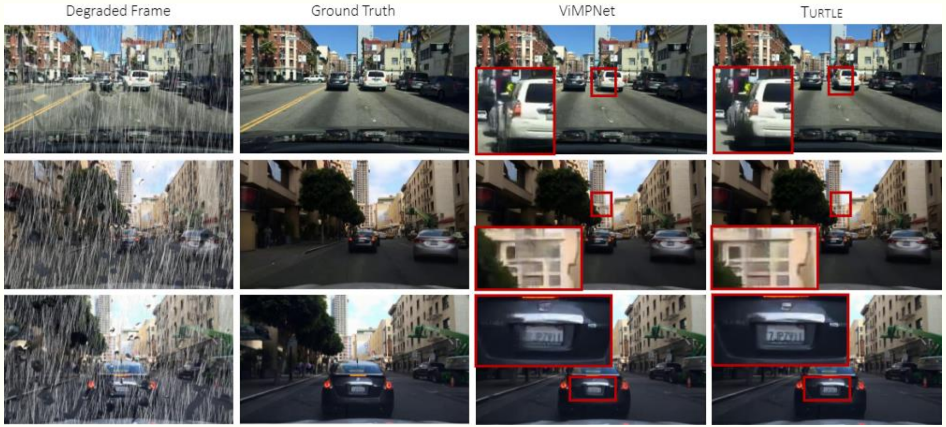


Figure 10: Video RainStreak and Raindrop Removal Results

Figure 11: Real-World Restoration Results (videos taken from a free videos website)

Any Questions?

ghasemab@ualberta.ca / mjanjua@ualberta.ca

https://github.com/Ascend-Research/Turtle

https://kjanjua26.github.io/turtle

Thanks!